

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application:

Listing of Claims:

1. (Currently amended) A method for clustering a plurality of items, each of the items including information, the method comprising:

inputting a plurality of items, each of the items including information, into a clustering process;

inputting an initial organization structure into the clustering process, the initial organization structure including one or more categories, a first one of the items being associated with at least one of the categories of the initial organization structure;

processing, using at least processing hardware, the plurality of items in at least the clustering process, the clustering process being based in part upon the initial organization structure and in part upon the information in each of the items, the clustering process being biased toward the categories of the initial organization structure, wherein the processing comprises determining a likeness level between a first item and a second item, the likeness level being defined such that the likeness level between two items is increased to the extent that both of the two items are associated with a same one or more of the categories of the initial organization structure, and wherein determining the likeness level between the first item and the second item comprises:

associating a first feature vector with the first item and a second feature vector with the second item, each feature vector representing information included in the corresponding item;

adding a first additional feature to each of the first feature vector and the second feature vector, the first additional feature representing a first category of the initial organization structure, the first additional feature in the first feature vector corresponding to a degree to which the first item is similar to one or more items associated with the first category of

the initial organization structure, the first additional feature in the second feature vector corresponding to a degree to which the second item is similar to one or more items associated with the first category of the initial organization structure; and

calculating a degree of similarity of the first item and the second item including calculating a similarity measure between the first additional feature in the first feature vector and the second feature vector;

automatically determining, using at least the processing hardware, a resulting organization structure based upon the processing of the plurality of items, the resulting organization structure comprising at least a portion of the initial organization structure and a modification to at least one of the categories of the initial organization structure such that an association of a first one of the items to at least one of the categories of the resulting organization structure is different from an association of the first one of the items to at least one of the categories of the initial organization structure; and

storing the resulting organization structure in memory.

2 - 4. (Canceled)

5. (Original) The method of claim 1 wherein the resulting organization structure relates to the initial organization structure based upon a category similarity.

6. (Original) The method of claim 1 wherein the resulting organization structure relates to the initial organization structure based upon a similarity of hierarchy structure.

7. (Original) The method of claim 1 wherein the item is a document, a product, a person, a DNA sequence, a purchase transaction, a financial record, or a species description.

8. (Currently amended) The method of claim[[s]] 1 further comprising outputting the resulting organization structure on an output device.

9. (Original) The method of claim 1 wherein the processing hardware uses at least a 500 MHz clock to efficiently run the clustering process.

10. (Original) The method of claim 1 wherein the plurality of items includes at least 10,000 items.

11. (Currently amended) A computer aided information organization device, the device including a computer readable storage medium encoded with program code, the program code comprising: one or more computer memories, the one or more computer memories including:

a first code directed to inputting at least 10,000 items in electronic form into a clustering process, each of the items including information;

a second code directed to inputting an initial organization structure in electronic form into the clustering process, the initial organization structure including one or more categories, a first one of the items being associated with at least one of the categories of the initial organization structure;

a third code directed to processing, using at least processing hardware, the plurality of items in at least the clustering process, the clustering process being based in part upon the initial organization structure and in part upon the information in each of the items;

a fourth code directed to automatically determining, using at least the processing hardware, a resulting organization structure based upon the processing, the resulting organization structure comprising at least a portion of the initial organization structure and a modification to at least one of the categories of the initial organization structure such that an association of the first one of the items to at least one of the categories of the resulting organization structure is different from an association of the first one of the items to at least one of the categories of the initial organization structure; ~~and~~

a fifth code directed to storing the resulting organization structure in ~~the one or more memories or another memory; and~~

a sixth code directed to determining a likeness level between a first item and a second item, the likeness level being defined such that the likeness level between two items

increases if both of the two items are associated with a same one or more of the categories in the initial organization structure, and wherein determining the likeness level between the first item and the second item comprises:

a code directed to associating a first feature vector with the first item and a second feature vector with the second item, each feature vector relating to information included in the corresponding item;

a code directed to extending the feature vector of each of the first item and the second item with an additional feature representing a category of the initial organization structure, the additional feature in each feature vector relating to a degree to which respective item is similar to one or more items associated with the corresponding category of the initial organization structure; and

a code directed to calculating a measure of similarity of the first item with the second item including calculating the similarity measure between the additional feature in the first feature vector and the second feature vector.

12 - 14. (Canceled)

15. (Currently amended) The device of claim 11, wherein the program code further comprises ~~further comprising~~ a seventh ~~sixth~~ code directed to outputting the resulting organization structure, the resulting organization structure including a plurality of categories.

16. (Currently amended) The device of claim 15, wherein the program code further comprises ~~further comprising~~ an eighth ~~seventh~~ code directed to inputting additional items using the resulting organization structure.

17. (Currently amended) The device of claim 11, wherein the program code further comprises ~~further comprising~~ a seventh ~~sixth~~ code directed to independently modifying the resulting organization structure using a graphical user interface.

18. (Original) The device of claim 17 wherein the independently modifying is provided by a user coupled to the graphical user interface.

19. (Previously presented) A computer implemented method for clustering a plurality of items, the method comprising:

inputting a first hierarchy, the first hierarchy including at least a parent category and a first child category;

inputting a plurality of items, each of the plurality of items including information, at least one of the items being associated with one of the categories in the first hierarchy;

processing by the computer the plurality of items in a clustering process based upon at least the first hierarchy and the information in each of the items;

automatically determining by the computer a second hierarchy based upon the processing of the plurality of items in the clustering process, the second hierarchy including a portion of the first hierarchy and one or more additional categories coupled to the first hierarchy, at least one of the one or more additional categories being a second child category of the parent category of the first hierarchy;

storing the second hierarchy in memory; and

assigning each of the plurality of items to a category of the second hierarchy.

20. (Canceled)

21. (Currently amended) The method of claim 1 [[20]] wherein the modification to at least one of the categories in the initial organization structure includes a modification to a first category such that the first item is associated with the first category of the resulting organization structure but is not associated with the first category of the initial organization structure.

22. (Currently amended) The method of claim 1 [[20]] wherein the modification to at least one of the categories in the initial organization structure includes a modification to a first category such that the first item is associated with the first category of the initial organization structure but is not associated with the first category of the resulting organization structure.

23. (Previously presented) The method of claim 1 wherein the resulting organization structure includes at least one additional category coupled to the initial organization structure.

24. (Previously presented) The method of claim 19 wherein the clustering process guarantees that the resulting organization structure is an extension of the initial organization structure.

25. (Previously presented) The method of claim 24 wherein processing in the clustering process comprises:

- forming a tentative cluster from at least two of the items;
- determining whether the tentative cluster is consistent with an extension of the initial organization structure; and
- rejecting the tentative cluster in the event that the tentative cluster is inconsistent with an extension of the initial organization structure.

26. (Previously presented) The method of claim 19 wherein the first hierarchy includes at least a parent category and a child category and wherein the guided clustering process includes:

- interposing a new category between a parent category of the initial organization structure and a child category of the initial organization structure.

27 (Previously presented) A method for clustering a plurality of items, each of the items including information, the method comprising:

- inputting a plurality of items, each of the items including information, into a clustering process;

- inputting an initial organization structure into the clustering process, the initial organization structure including one or more categories, wherein at least a first one of the items is not associated with any of the categories of the initial organization structure;

processing, using at least processing hardware, the plurality of items in at least the clustering process, the clustering process being based in part upon the initial organization structure and in part upon the information in each of the items;

automatically determining, using at least the processing hardware, a resulting organization structure based upon the processing of the plurality of items, the resulting organization structure comprising at least a portion of the initial organization structure and a modification to the initial organization structure such that the first one of the items is associated with at least one of the categories in the resulting organization structure; and

storing the resulting organization structure in memory.